JURNAL GAUSSIAN, Volume 1, Nomor 1, Tahun 2012, Halaman 83-92
Online di: http://ejournal-s1.undip.ac.id/index.php/gaussian

MODEL REGRESI DATA TAHAN HIDUP TERSENSOR TIPE III BERDISTRIBUSI LOG-LOGISTIK

Ibnu Athoillah[1], Triastuti Wuryandari[2], Sudarno[3]
[1]Mahasiswa Jurusan Statistika FSM Universitas Diponegoro
[2,3]Staf Pengajar Jurusan Statistika FSM UNDIP

ABSTRAK

Waktu hidup T merupakan waktu dari awal perlakuan sampai terjadinya respon pertama kali yang ingin diamati yang dapat berupa kematian karena penyakit tertentu, keadaan sakit yang terulang kembali setelah pengobatan atau munculnya penyakit baru. Dalam penelitian uji ketahanan hidup terdapat istilah data tersensor dan tidak tersensor. Pengamatan tersensor terjadi jika waktu tahan hidup dari individu yang diamati tidak diketahui secara pasti sedangkan pengamatan tidak tersensor merupakan pengamatan waktu tahan hidup yang diketahui dengan pasti. Ada 3 macam jenis penyensoran dalam pengamatan terhadap waktu tahan hidup yaitu penyensoran tipe I, penyensoran tipe II dan penyensoran tipe III. Penyensoran tipe III merupakan suatu pengamatan yang dilakukan terhadap beberapa individu pada waktu yang berlainan dalam jangka waktu tertentu, hal ini dikarenakan suatu individu masuk ke dalam pengamatan pada waktu yang berbeda. Pengaruh faktor lain terhadap variabel respon yang berupa waktu tahan hidup patut dipertimbangkan hubungannya. Salah satu cara untuk mengetahui hubungannya adalah melalui model regresi. Model regresi data tahan hidup tersensor tipe III berdistribusi Log-logistik dibuat mengikuti bentuk distribusi variabel responnya. Estimasi parameter yang digunakan metode maksimum likelihood. Model regresi ini diaplikasikan untuk mengestimasi waktu tahan hidup pasien kanker paru-paru terhadap faktor jenis sel yang terinfeksi dan jenis perawatan.

Kata kunci : Regresi, Data Tahan Hidup, Distribusi Log-Logistik, Data Tersensor Tipe III

ABSTRACT

Lifetime T is the time from initial treatment until the first response is to be observed which can be death due to a particular disease, illness that recur after treatment or the emergence of new diseases. In research of survival testing the data term are censored and not censored. Censored observation occur if the survival time of the observed individual is not known with certainty while the observation not censored if the survival time of observation is known with certainty. There are three different types of censoring observation that are type I,type II, and type III. Censoring type III is an observation made to several individuals at different time within a certain period, this is because an individual entry into the observations at different times. Influence of other factros on the response variable that is survival time relation should be considered. One way to know relationship is through a regression model. Regression model of survival data with censoring type III of log-logistic distribution is made following the curve of the response variable. Estimation of parameters using maximum likelihood methods. Regression model was apllied to estimate the survival time of patients with lung cancer for factors of the infected cell and type of treatment.

Keyword : Regression, Survival Data, Log-Logistic Distribution, Type III of Censored data

1. PENDAHULUAN

Dalam penelitian uji ketahanan hidup, satu hal yang menarik adalah data waktu hidup dapat berupa data tersensor dan tidak tersensor. Pengamatan tersensor terjadi jika waktu tahan hidup dari individu yang diamati tidak diketahui secara pasti. Pengamatan tersensor diindikasikan adanya individu yang tetap hidup sampai jangka waktu yang telah ditentukan sedangkan pengamatan tidak tersensor merupakan pengamatan jika semua individu atau unit data yang diteliti mati atau gagal sehingga waktu tahan hidupnya diketahui dengan pasti.

Ada 3 macam jenis penyensoran dalam pengamatan terhadap waktu tahan hidup yaitu sensor tipe I, sensor tipe II dan sensor tipe III. sensor tipe I dibatasi oleh waktu dan setiap individu masuk kedalam pengamatan secara bersama. Sensor tipe II dibatasi oleh banyaknya r individu yang gagal dari sebanyak n individu dalam pengamatan ($1 \leq r \leq n$). Sensor tipe III dibatasi oleh waktu dan setiap individu masuk kedalam pengamatan dalam waktu yang berbeda. Jenis penyensoran yang sering digunakan untuk mendeteksi waktu tahan hidup pasien adalah sensor tipe III karena beberapa pasien sangat jarang masuk ke dalam pengamatan dalam waktu yang sama..

Salah satu distribusi waktu ketahanan hidup adalah distribusi Log-Logistik. Distribusi Log-Logistik mempunyai bentuk fungsi kegagalan (hazard) yang tidak monoton naik atau monoton turun dan juga tidak konstan seperti pada distribusi Weibull dan Eksponensial. Misalnya dalam kasus transplantasi hati, beberapa hari pertama setelah transplantasi, tingkat kegagalan akan tinggi sampai mencapai puncak dan setelah tubuh mengalami penyesuaian selama beberapa hari maka tingkat kegagalan pun akan menurun.

Data tahan hidup dari beberapa individu dalam suatu observasi dapat dikembangkan dengan analisis regresi untuk menganalisis faktor–faktor atau variabel-variabel penjelas yang dapat mempengaruhi data tahan hidupnya seperti karakteristik individu, jenis perawatan, dan kondisi lingkungan.

2. TINJAUAN PUSTAKA

2.1 Konsep Dasar Uji Tahan Hidup

Waktu tahan hidup T adalah variabel random non negatif yang mewakili ketahanan hidup dari individu-individu dalam suatu populasi yang homogen. Distribusi probabilitas dari T dapat dispesifikasikan dalam banyak hal, tiga diantaranya sebagai fungsi dasar dalam aplikasi tahan hidup yaitu fungsi padat peluang, fungsi tahan hidup dan fungsi kegagalan (Kalbfleisch, 2002).

2.1.1 Fungsi padat peluang

Fungsi padat peluang adalah probabilitas suatu individu mati atau gagal dalam interval waktu dari t sampai $t + \Delta t$, dengan T merupakan variabel random. Fungsi padat peluang dinyatakan dengan

$$f(t) = \lim_{\Delta t \to 0} [\frac{P(objek\ gagal\ pada\ interval\ t, (t + \Delta t))}{\Delta t}]$$

$$= \lim_{\Delta t \to 0} \left[\frac{P(\ t < T < t + \Delta t)}{\Delta t}\right]$$

2.1.2 Fungsi tahan hidup

Fungsi tahan hidup adalah probabilitas suatu individu yang masih dapat bertahan hidup lebih dari waktu *t*. Jika T sebagai variabel random waktu tahan hidup dalam interval $[0,\infty)$, maka $S(t)$ dapat dirumuskan

$$S(t) = P(objek\ hidup\ lebih\ dari\ waktu\ t)$$

$$= P(T > t)$$

$$= 1 - P(objek\ gagal\ sebelum\ waktu\ t)$$

$$= 1 - P(T \leq t)$$

2.1.3 Fungsi kegagalan

Fungsi kegagalan atau fungsi hazard menyatakan peluang kegagalan suatu individu pada waktu *t*, jika diketahui bahwa individu tersebut tetap hidup hingga waktu *t*. Fungsi kegagalan dari waktu tahan hidup T dinotasikan dengan

$$h(t) = \lim_{\Delta t \to 0} \left[\frac{P\ (t<T<t+\Delta t | T>t)}{\Delta t} \right]$$

(Lee E.T, 2003).

2.2 Data Tersensor

Ada tiga tipe penyensoran yang sering digunakan dalam eksperimen uji hidup, yaitu sebagai berikut:

1. Sensor tipe I

   Sensor tipe I adalah tipe penyensoran dimana percobaan akan dihentikan setelah mencapai waktu T yang telah ditentukan untuk mengakhiri semua n individu yang masuk pada waktu yang sama.

2. Sensor tipe II

   Sensor tipe II adalah tipe penyensoran dimana sampel ke – r merupakan observasi terkecil dalam sampel random berukuran $n(1 \leq r \leq n)$.

3. Sensor tipe III

   Pada sensor tipe III individu atau unit uji masuk ke dalam percobaan pada waktu yang berlainan selama periode waktu tertentu.

(Lee E.T, 2003).

2.3 Maksimum Likelihood Estimation (MLE)

Metode untuk mengestimasi nilai parameter distribusi dari data dalam fungsi tahan hidup adalah dengan menggunakan metode maksimum likelihood. Misalkan $X_1, X_2, \ldots X_n$ adalah sampel random dari populasi dengan fungsi padat peluang $f(x|\theta_1, \theta_2, \ldots, \theta_m)$, fungsi likelihood didefinisikan dengan

$$L(\theta_1, \theta_2, \ldots, \theta_m | X) = \prod_{i=1}^{n} f(x_i | \theta_1, \theta_2, \ldots, \theta_m) \qquad (2.12)$$

Bila fungsi likelihood terdeferensialkan dalam $\theta_1, \theta_2, \ldots, \theta_m$, maka calon estimator maksimum likelihood yang mungkin adalah harga-harga $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m$ sedemikian sehingga

$$\frac{\partial}{\partial \theta_k} L(\theta_1, \theta_2, \ldots, \theta_m | X) = 0 \qquad k = 1,2,\ldots,m \qquad (2.13)$$

Untuk membuktikan bahwa $\hat{\theta}_k$ benar – benar memaksimumkan fungsi likelihood, $L(\theta_1, \theta_2, \ldots, \theta_m | X)$ harus ditunjukkan bahwa

$$\frac{\partial^2}{\partial {\theta_k}^2} L(\theta_1, \theta_2, \ldots, \theta_m | X) | \theta_k < 0 \qquad (2.14)$$

Dalam banyak kasus, diferensi digunakan pada logaritma natural dari $L(\theta_1, \theta_2, \ldots, \theta_m | X)$, yaitu $l = \log L(\theta_1, \theta_2, \ldots, \theta_m | X)$. Hal ini dimungkinkan karena fungsi log naik tegas pada $(0, \infty)$, yang berarti bahwa $L(\theta_1, \theta_2, \ldots, \theta_m | X)$ dan $l = \log L\,(\theta_1, \theta_2, \ldots, \theta_m | X)$ mempunyai ekstrim yang sama .

(Bain, L.J and Engelhart, M. 1992)

2.4 Distribusi Log-Logistik

Distribusi Log-Logistik ini berasal dari distribusi Logistik dengan variabel random *Y* yang mempunyai fungsi densitas peluang

$$f(y) = \frac{\sigma^{-1}\exp\left[\frac{(y-\mu)}{\sigma}\right]}{\{1+\exp\left[\frac{y-\mu}{\sigma}\right]\}^2}, -\infty < y < \infty \tag{1}$$

Dengan $\mu = log\ \alpha$, $\sigma = \beta^{-1}$, a $-\infty < \mu < \infty$ dan $\sigma > 0$.

Variabel random T dikatakan mengikuti distribusi Log-Logistik dengan parameter α, β jika mempunyai fungsi densitas :

$$f(t) = \frac{\left(\frac{\beta}{\alpha}\right)(\frac{t}{\alpha})^{\beta-1}}{[1+(\frac{t}{\alpha})^{\beta}]^2}, t > 0, \alpha > 0, \beta > 0 \tag{2}$$

Fungsi distribusi kumulatifnya adalah

$$F(t) = \frac{t^{\beta}}{\alpha^{\beta}+t^{\beta}}, t > 0, \alpha > 0, \beta > 0$$

Fungsi tahan hidupnya adalah

$$S(t) = \frac{1}{1+(\frac{t}{\alpha})^{\beta}}, t > 0, \alpha > 0, \beta > 0 \tag{3}$$

Fungsi kegagalannya adalah

$$h(t) = \frac{\left(\frac{\beta}{\alpha}\right)(\frac{t}{\alpha})^{\beta-1}}{1+(\frac{t}{\alpha})^{\beta}}, t > 0, \alpha > 0, \beta > 0$$

2.5 Estimasi Parameter Data Tersensor Tipe III

Data tersensor tipe III diasumsikan memiliki waktu tahan hidup T dan waktu tersensor L, dengan T dan L merupakan variabel random kontinu dengan fungsi ketahanan *S(t)* dan *G(t)*. Sampel tersensor tipe III muncul ketika individu 1,2 ,..., n dibatasi oleh waktu pengamatan $L_1, L_2, \dots, L_n$, jadi waktu tahan hidup suatu individu T_i hanya teramati jika $T_i \leq L_i$. Saat (T_i, L_i) , $i = 1,2, \dots, n$ independen maka

$$t_i = \min(T_i, L_i) \text{ dan } \delta_i = \begin{cases} 1\ jika\ T_i \leq L_i \\ 0\ jika\ T_i > L_i \end{cases}$$

$$L = \prod_1^n [f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}]$$

2.6 Uji Kecocokan Distribusi Log-Logistik

Jika T adalah variabel random dengan distribusi kontinu $F(t)$, dan mempertimbangkan rumusan hipotesis

$H_0: F(t) = F_0(t)$ (Data mengikuti distribusi Log-Logistik)
$H_1: F(t) \neq F_0(t)$ (Data tidak mengikuti distribusi Log-Logistik)

Statistik Anderson Darling untuk data tersensor tipe III adalah sebagai berikut :

$$A_{n,p}^2 = -\sum_{i=1}^{r}\left(\frac{2i-1}{n}\log F_0\left(t_{(i)}\right) - \frac{2n-2i+1}{n}\log[1-F_0\left(t_{(i)}\right)]\right)$$
$$+\frac{r^2}{n}\log F_0(L) - \frac{(n-r)^2}{n}\log[1-\mathrm{F}_0(\mathrm{L})] - nF_0(L)$$

Koziol dan Byar (1975) menetapkan bahwa $A_{n,p}^2$ mendekati distribusi dari $D_{n,p}$ dalam kasus data tersensor tipe III , p=$F_0(L)$.

Kriteria penolakan : H_0 ditolak jika $A_{n,p}^2 > D_{n,p}$ (Lawless , 1982)

2.7 Regresi Variabel Dummy

Analisis regresi variabel dummy merupakan analisis regresi dengan variabel independen kualitatif. Variabel independen kualitatif tersebut dapat berupa kelas, kelompok atau tingkatan. Menggunakan pola koding biner (0,1), variabel dummy selalu berbentuk dikotomi. Semua responden yang menjadi anggota kategori diberi kode 1 sedangkan responden yang tidak dalam kategori tersebut diberi kode 0. Persamaan regresi variabel dummy adalah sebagai berikut :

$$Y_i = \beta_0 + \beta_1 d_{1i} + .. + \beta_m d_{mi} + \varepsilon_i, \qquad k = 1 \dots m$$

d_{ki} adalah variabel dummy, dengan nilai 0 atau 1.

3. MODEL REGRESI DATA TAHAN BERDISTRIBUSI LOG-LOGISTIK

Model regresi waktu tahan hidup dibentuk dalam model skala-lokasi. Dalam model ini, waktu tahan hidup T ditransformasikan logaritma sehingga menjadi Y=log t dan diperoleh persamaan regresi

$$Y_i = \mu(x_i) + \sigma\varepsilon_i$$

dengan $\sigma > 0$ dan ε memiliki distribusi Logistik standar (Lawless , 1982).

Dalam model regresi Log-Logistik, koefisien *slope* dapat dinyatakan sebagaimana Odd Rasio. Dalam hal membentuk Odd Rasio, pertama dinyatakan terlebih dahulu fungsi tahan hidup sebagai peluang sukses dari faktor yang dipertimbangkan yaitu

$$S(t_i, x_i, \alpha \;\;, \beta) = (1 + \exp(\varepsilon_i))^{-1}$$

Dengan ε_i = $(Y_i - \mu(x_i))/\sigma$, Y_i = log t_i dan $\mu(x_i) = \theta_0 + \sum_{k=1}^{m}\theta_k x_{ki}$ sehingga sesuai dengan persamaan dengan $\mu(x_i) = \log\alpha\,(x_i)$ dan $\sigma = \beta^{-1}$

Maka Odd dari waktu ketahanan *t* adalah

$$\frac{S(t_i,x_i,\alpha \;\;,\beta)}{1-S(t_i,x_i,\alpha \;\;,\beta)} = \exp\,(-\varepsilon_i)$$

Jika Odd tersebut di log kan maka akan menjadi

$$\log\frac{S(t_i,x_i,\alpha \;\;,\beta)}{1-S(t_i,x_i,\alpha \;\;,\beta)} = -\log t_i/\sigma + \frac{\theta_0}{\sigma} + \frac{1}{\sigma}\sum_{k=1}^{m}\theta_k x_{ki}$$

Misalkan OR_i dan OR_j adalah Odd dari waktu ketahanan *t* untuk individu *i* dan *j*, maka logaritma dari Odd Rasio adalah

$$\log\frac{OR_i}{OR_j} = \frac{1}{\sigma}\sum_{k=1}^{m} a_k\,(x_{ki} - x_{kj})$$

Rasio ini independen terhadap waktu *t* sehingga individu yang berbeda memiliki odd rasio yang sama pula. Oleh karena itu, model regresi Log-Logistik adalah model proporsional odds bukan model proporsional hazard.

3.1 Model Skala Lokasi untuk Log t

Diketahui T adalah variabel random berdistribusi Log-Logistik dengan fungsi padat peluang T yang diberikan oleh **x** tertentu. Jika T ditransformasikan dengan transformasi $Y = \log t$ maka diperoleh fungsi padat peluang Y yang diberikan oleh **x** tertentu yang identik dengan pesamaan (1) yaitu

$$f(y|\boldsymbol{x}) = \frac{\exp\left(\frac{y-\mu(x)}{\sigma}\right)}{\sigma\left[1+\exp\left(\frac{y-\mu(x)}{\sigma}\right)\right]^2}, -\infty < y < \infty \qquad (4)$$

Fungsi distribusi kumulatifnya adalah

$$F(y|\boldsymbol{x}) = 1 - \frac{1}{1+\exp\left(\frac{y-\mu(x)}{\sigma}\right)}$$

Fungsi tahan hidupnya adalah adalah

$$S(y|\boldsymbol{x}) = \frac{1}{1+\exp\left(\frac{y-\mu(x)}{\sigma}\right)} \qquad (5)$$

Fungsi kegagalannya adalah

$$h(y|\boldsymbol{x}) = \frac{\exp\left(\frac{y-\mu(x)}{\sigma}\right)}{\sigma\left[1+\exp\left(\frac{y-\mu(x)}{\sigma}\right)\right]}$$

3.2 Estimasi Titik

Diketahui fungsi likelihood adalah

$$L = \prod_1^n f(t_i)^{\delta_i}[S(t_i)]^{1-\delta_i}$$

Diketahui bahwa fungsi densitas dan fungsi tahan hidup seperti pada persamaan (2) dan (3) maka fungsi likelihoodnya adalah

$$L(\alpha,\beta) = \prod_{i=1}^n \left[\frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{t_i}{\alpha}\right)^{\beta-1}}{[1+\left(\frac{t_i}{\alpha}\right)^{\beta}]^2}\right]^{\delta_i}\left[\frac{1}{1+\left(\frac{t_i}{\alpha}\right)^{\beta}}\right]^{1-\delta_i} \qquad (6)$$

Dalam mengestimasi titik dari parameter distribusi Log-Logistik dapat dilakukan dengan menurunkan persamaan (6) terhadap α dan β *kemudian* disamadenganlan nol menjadi

$$\sum \delta_i - \sum_{i=1}^n \frac{(\delta_i+1)\left(\frac{t_i}{\alpha}\right)^{\beta}}{1+\left(\frac{t_i}{\alpha}\right)^{\beta}} = 0$$

dan

$$\frac{\sum \delta_i}{\beta} + \sum_{i=1}^n \delta_i \log\left(\frac{t_i}{\alpha}\right) - \sum_{i=1}^n \frac{(\delta_i+1)\left(\frac{t_i}{\alpha}\right)^{\beta}\log\left(\frac{t_i}{\alpha}\right)}{1+\left(\frac{t_i}{\alpha}\right)^{\beta}} = 0$$

karena kedua persamaan tidak dapat diselesaikan secara langsung, diperlukan suatu metode numerik untuk mendapatkan hasil estimasinya. Solusinya antara lain dengan menggunakan metode iteratif yaitu Newton-Raphson

3.3 Estimasi koefisien regresi

Misalkan tiap individu memiliki waktu tahan hidup t_i dan vektor regresi ***$x_{ik}=(x_{i1},\ldots,x_{im})$***. Log waktu tahan hidup Y yang mempunyai fungsi padat peluang dan fungsi tahan hidup yang diberikan oleh **x** tertentu berturut – turut (4) dan (6). Fungsi likelihood untuk sampel tersensor tipe III yang didasarkan pada n individu adalah

$$L(\theta) = \prod_{i=1}^n f(t_i)^{\delta_i}\,[S(t_i)]^{1-\delta_i}$$

$$L(\theta) = \prod_{i=1}^n \left[\frac{\exp\left(\frac{y_i-\mu(x)}{\sigma}\right)}{\sigma\left[1+\exp\left(\frac{y_i-\mu(x)}{\sigma}\right)\right]^2}\right]^{\delta_i}\left[\frac{1}{1+\exp\left(\frac{y_i-\mu(x)}{\sigma}\right)}\right]^{1-\delta_i}$$ Dengan $\mu(x) = \theta_0 + \sum_{k=1}^m \theta_k x_k = x\theta$

Misal $z_i = \frac{y_i - x\theta}{\sigma}$, maka turunan pertama dan kedua fungsi likelihood adalah

(i) $\frac{\partial log\ l}{\partial \theta_l} = \sum_{i=1}^{n} - \frac{\delta_i x_{il}}{\sigma} + \sum_{i=1}^{n} \frac{(\delta_i+1)\left(\frac{x_{il}}{\sigma}\right)\exp(z_i)}{1+\exp(z_i)}$

(ii) $\frac{\partial log\ l}{\partial \sigma} = -\frac{1}{\sigma}\sum_{i=1}^{n}\left(\delta_i(z_i)\right) - \sum \frac{\delta_i}{\sigma} + \sum_{i=1}^{n} \frac{(\delta_i+1)\left(\frac{z_i}{\sigma}\right)\exp(z_i)}{1+\exp(z_i)}$

(iii) $\frac{\partial log\ l}{\partial \theta_l \partial \theta_s} = -\sum_{i=1}^{n} \frac{(\delta_i+1)}{\sigma^2} x_{il} x_{is} \left[\frac{[\exp(z_i)][1+\exp(z_i)] - [[\exp(z_i)]]^2}{[1+\exp(z_i)]^2}\right]$

(iv) $\frac{\partial log\ l}{\partial \sigma^2} = \sum_{i=1}^{n} \frac{2\left(\delta_i(z_i)\right)}{\sigma^2} + \sum \frac{\delta_i}{\sigma^2} + \sum_{i=1}^{n} \frac{(\delta_i+1)\left[\left(\frac{z_i}{\sigma^2}\exp(z_i)[z_i-2]\right)(1+\exp(z_i))\right] - \left[\left(\frac{-1}{\sigma}z_i\exp(z_i)\right)\left(\left(\frac{z_i}{\sigma}\right)\exp(z_i)\right)\right]}{\left(1+\exp(z_i)\right)^2}$

(v) $\frac{\partial log\ l}{\partial \theta_l \partial \sigma} = \sum_{i=1}^{n} \frac{\delta_i x_{il}}{\sigma^2} + \sum_{i=1}^{n} \frac{(\delta_i+1)\left(\frac{x_{il}}{\sigma^2}\exp(z_i)[z_i-1]\right)(1+\exp(z_i)) - \left(\frac{-1}{\sigma}z_i\exp(z_i)\right)\left(\left(\frac{x_{il}}{\sigma}\right)\exp(z_i)\right)}{\left(1+\exp(z_i)\right)^2}$

(7)

Persamaan maksimum likelihood 3.18 dapat diselesaikan dengan metode iterasi Newton-Raphson dengan langkah-langkah seperti pada estimasi titik.

4. STUDI KASUS PADA PASIEN KANKER PARU-PARU

4.1 Data

Jenis data yang digunakan adalah data sekunder dari data data ketahanan pasien kanker paru-paru pasca operasi berjumlah 97 pasien. Data diperleh dari *The Statistical Analysis of Failure Time Data. 2nd Edition.2002. John Wiley& Sons, Inc*

Variabel dependennya adalah waktu ketahanan pasien pasca operasi.. Statusnya adalah apakah waktu ketahanan ini diketahui pasti atau tersensor Variabel independen terdiri dari 2 variabel yaitu jenis sel yang terinfeksi yang terdiri dari empat tipe yaitu *sel squamos, sel smalll, sel adeno, sel large* dan jenis perawatan yang terdiri dari 2 jenis yaitu standar dan non standar.

Data yang diperoleh dianalisis menggunakan software minitab 14 dengan mengasumsikan bahwa model regresi yang terbentuk berdistribusi Log-Logistik dengan error berdistribusi Logistik Standar.

4.2 Pengolahan Data

pengolahan data untuk menentukan model regresi data tahan hidup tersensor tipe III berdistribusi Log-Logistik, dilakukan dengan langkah-langkah berikut ini :

1. Memilih data tahan hidup tersensor tipe III
2. Menguji kecocokan data apakah mengikuti distribusi Log-Logistik
3. Mengestimasi parameter dengan metode maksimum likelihood
4. Membentuk model regresi data tahan hidup
5. Menguji signifikansi koefisien regresi secara simultan.
6. Menguji signifikansi koefisien regresi secara parsial.
7. Menentukan model regresi terbaik.

4.3 Hasil dan Analisis

4.3.1 Uji kecocokan distribusi Log-Logistik

Dengan menggunakan metode maksimum likelihood estimation dalam software Minitab didapat nilai $\mu = 4.24032$; $\sigma = 0.722982$; maka

$\alpha = \exp(4.24032) = 69.4301$ dan

$\beta = 1/0.722982 = 1.3832$

Hasil uji kecocokan distribusi Log-Logistik untuk data pasien kanker paru-paru adalah:
Hipotesis :

$H_0: F(t) = F_0(t)$ (Data berdistribusi Log-Logistik)

$H_1: F(t) \neq F_0(t)$ (Data tidak berdistribusi Log-Logistik)

Statistik Uji :

$$A_{n,p}^2 = -\sum_{i=1}^{r}\left(\frac{2i-1}{n}\log F_0\left(t_{(i)}\right) - \frac{2n-2i+1}{n}\log[1-F_0\left(t_{(i)}\right)]\right) + \frac{r^2}{n}\log F_0(L) - \frac{(n-r)^2}{n}\log[1-\mathrm{F}_0(\mathrm{L})] - nF_0(L)$$

Kriteria penolakan :
Tolak H_0 jika H_0 ditolak jika $A_{n,p}^2 > D_{n,p}^{\alpha}$
Keputusan :
Berdasarkan hasil output Minitab didapat nilai uji Anderson Darling sebesar 0.563.dan $p = F_0(L) = 0.949852 \approx 0{,}9$, maka tabel Koizol dan Byar , $D_{n,p}^{\alpha} = D_{97;0,9}^{5\%} = 1.3581$. Diketahui $A_{97,0.9}^2 = 0{,}563 < D_{97;0,9}^{5\%} = 1.3581$, sehingga H_0 diterima yang berarti bahwa data waktu ketahanan pasien kanker paru-paru berdistribusi Log-Logistik.

4.3.2 Membentuk model regresi data tahan hidup

Dalam membentuk model regresi data tahan hidup dimasukkan semua variabel independen ke dalam model regresi, didapatkan hasil sebagai berikut:

Tabel 1 Hasil Regresi Waktu Tahan Hidup Berdistribusi Log-Logistik

Variabel	Koefisien	P	Z(hitung)
Intercept	4.66932	0.000	16.12
Squamos (X_{11})	-0.00309	0.993	-0.01
Small (X_{12})	-1.15106	0.000	-3.70
Adeno (X_{13})	-0.821444	0.017	-2.38
Perawatan (X_2)	0.413446	0.076	1.78
Scale	0.637055		

Model ini menggambarkan hubungan antara waktu tahan hidup terhadap tipe sel yang terinfeksi kanker dan variabel perawatan yang dijelaskan melalui variabel dummy.
Model regresinya adalah :
t_i=exp(4.66932-0.00309x_{11i}-1.15106x_{12i}-0.821444x_{13i}+0.413446x_{2i}+0.637055 ε_i)
Dengan menggunakan model $Y = log\ t$ persamaanya menjadi :
y_i = 4.66932 - 0.00309x_{11i} - 1.15106x_{12i} - 0.821444x_{13i} + 0.413446x_{2i} +0.637055 ε_i

4.3.3 Uji koefisien regresi secara simultan

Hipotesis :

$H_0: \theta_1 = \theta_2 =..= \theta_4 = 0$ untuk semua k dengan $k= 1,\ldots 4$
(Secara simultan, koefisien regresi tidak signifikan secara statistik)

$H_1: \theta_k \neq 0$ untuk paling sedikit satu k dengan $k= 1,\ldots 4$
(Secara simultan, koefisien regresi signifikan secara statistik)

Taraf Signifikansi $\alpha = 5\%$

Statistik uji $\Lambda = -2\log\left(\frac{L(\tilde{\theta},\tilde{\sigma})}{L(\widehat{\theta,\hat{\sigma}})}\right)$:

$= -2(-536.124 + 517.868)$

$= 36.512$

Kriteria Penolakan : Tolak H_0 jika P-value < 5 % atau $\Lambda > \chi_4^2$

Keputusan : H_0 ditolak karena nilai $\Lambda > \chi^2_{0.05;\,4}$ yaitu 36.512 > 9.49

Kesimpulan : Karena H_0 ditolak maka dapat disimpulkan bahwa koefisien regresi secara simultan signifikan.

4.3.4 Uji koefisien regresi secara parsial

Berdasarkan output Minitab 14 diperoleh hasil untuk masing-masing koefisien regresi sebagai berikut:

Tabel 2 Koefisien Regresi Data Tahan Hidup Tersensor Tipe II Berdistribusi Log-Logistik

		Standard			95.0% Normal CI	
Predictor	Coef	Error	Z	P	Lower	Upper
Intercept	4.66932	0.289679	16.12	0.000	4.10156	5.23708
Squamos	-0.00309	0.366050	-0.01	0.993	-0.720525	0.714355
Small	-1.15106	0.310833	-3.70	0.000	-1.76028	-0.541837
Adeno	-0.821444	0.345221	-2.38	0.017	-1.49807	-0.144823
Perawatan	0.413446	0.232882	1.78	0.076	-0.0429956	0.869887
Scale	0.637055	0.055675			0.536768	0.756079

Hipotesis :

$H_0: \theta_k = 0$ untuk suatu k dengan $k= 1, \ldots 4$

(secara parsial, koefisien regresi tidak signifikan secara statistik)

$H_1: \theta_k \neq 0$ untuk suatu i dengan $k= 1, \ldots 4$

(secara parsial, koefisien regresi signifikan secara statistik)

Taraf Signifikansi $\alpha = 5\%$

Statistik uji : $Z_{hitung} = (\hat{\theta}_k - 0)/se(\hat{\theta}_k)$

Kriteria Penolakan : Tolak H_0 jika P-value < 5 % atau $|Z_{hitung}| > Z_{tabel}$ dimana $Z_{tabel} = Z_{\frac{\alpha}{2}}$

$$\theta_1 \rightarrow |Z_{hitung}| = 0.01, P - value = 0.993$$
$$\theta_2 \rightarrow |Z_{hitung}| = 3.70, P - value = 0.000$$
$$\theta_3 \rightarrow |Z_{hitung}| = 2.38, P - value = 0.017$$
$$\theta_4 \rightarrow |Z_{hitung}| = 1.78, P - value = 0.076$$

Keputusan

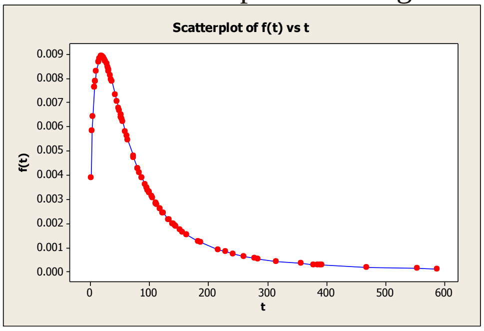
H_0 ditolak pada koefisien yang memiliki P-value < 5 % atau $|Z_{hitung}| > Z_{tabel} = 1.96$,sehingga koefisien θ_2 dan $\theta_3$adalah signifikan sedangkan koefisien regresi θ_1 dan θ_4 tidak signifikan.

Kesimpulan

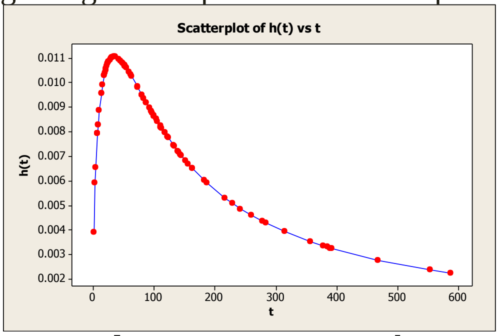
Hanya koefisien regresi θ_2 dan θ_3 yang signifikan dalam model.

4.3.5 Analisa grafik

berikut ini ditampilkan hasil grafik dari tiga fungsi konsep dasar tahan hidup



Gambar 1. Fungsi Padat Peluang



Gambar 2. Fungsi Kegagalan

Berdasarkan gambar 1 dapat dijelaskan bahwa waktu tahan hidup seorang pasien yang rendah memiliki peluang untuk gagal yang juga rendah. Semakin tinggi waktu tahan hidup seorang pasien tersebut semakin tinggi pula peluang untuk gagal sampai mencapai waktu puncak. Dalam kasus ini waktu puncaknya saat t = 19, pada ssat t = 19 peluang gagal mencapai titik tertinggi. Setelah melewati waktu tersebut, waktu tahan hidup seorang pasien yang semakin tinggi justru memiliki peluang untuk gagal yang semakin rendah.

Berdasarkan gambar 2 dapat dijelaskan bahwa waktu tahan hidup seorang pasien yang rendah memiliki tingkat kegagalan yang rendah, semakin tinggi waktu tahan hidup dari seorang pasien, maka semakin tinggi pula tingkat kegagalan pasien tersebut sampai mencapai titik puncak dengan tingkat kegagalan tertinggi, diketahui waktu tahan hidup t mencapai titik puncak pada $t= 35$ dan setelah melewati waktu tersebut maka waktu ketahanan yang semakin tinggi,tingkat kegagalan dari seorang pasien akan semakin menurun

5. KESIMPULAN

Model regresi data tahan hidup tersensor tipe III berdistribusi Log-logistik adalah $Y_i = \mu(x_i) + \sigma\varepsilon_i$, dengan $\mu(x_i) = \theta_0 + \sum_{k=1}^{m} \theta_k x_{ki}$, dan error berdistribusi logistik standar. Dengan

$$U(\hat{\theta}_k) = \sum_{i=1}^{n} -\frac{\delta_i x_{il}}{\sigma} + \sum_{i=1}^{n} \frac{(\delta_i+1)\left(\frac{x_{il}}{\sigma}\right)\exp(z_i)}{1+\exp(z_i)}, \text{ dan}$$

$$U(\hat{\sigma}) = -\frac{1}{\sigma}\sum_{i=1}^{n}\left(\delta_i(z_i)\right) - \sum_{i=1}^{n}\frac{\delta_i}{\sigma} + \sum_{i=1}^{n}\frac{(\delta_i+1)\left(\frac{z_i}{\sigma}\right)\exp(z_i)}{1+\exp(z_i)}$$

kedua persamaan diatas harus diselesaikan dengan metode itersi Newton-Raphson.

Dari contoh penerapan kasus pada pasien kanker paru-paru dengan variabel independen jenis sel terinfeksi dan jenis perawatan, model regresi yang terbentuk adalah

$$y_i = 4.66932 - 0.00309x_{11i} - 1.15106x_{12i} - 0.821444x_{13i} + 0.413446x_{2i} + 0.637055\,\varepsilon_i$$

DAFTAR PUSTAKA

Alakus, K. and Erilli, N.A. 2001. *Confidence Intervals Estimation for Survival Function in Log-logistic Distribution and Proportional Odds Regression Based on Censored Survival Time Data.* J Biomet Biostat, 2:116.

Bain, L.J. and Engelhart, M. 1992. *Introduction to Promodulity and Mathematical Statistics, 2nd ed.* Duxbury Press, Belmont. California.

Bennet, S. 1983. *Log-logistic Regression Models for Survival Data*. Applied Statistics, 32. 165-171.

James, H.S. 2008. *Models For probability and Statistical Inference: Theory and Application*. John Willey & Sons, Inc. Canada

Kalbfleisch and Prentice. 2002. *The Statistical Analysis of Failure Time Data 2nd ed.* Wiley & Sons, Inc. Canada.

Lawless, J.F. 1982. *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons, Inc. Canada.

Lawless, J.F. 2003. *Statistical Models and Methods for Lifetime Data 2nd Edition*. John Wiley & Sons, Inc. Canada.

Lee, E.T. 2003. *Statistical Methods for Survival Data Analysis 3rd Edition*. John Wiley & Sons, Inc . Canada.

Walpole, R.E. and Myers, R.H. 2007. *Probability and Statistic for Engineers and Scientist*. Prentice Hall International. New Jersey.